MILESTONE 2: UPDATE PROJECT PROGRESS

SMART WATER MANAGEMENT ASSISTANT

WITH CLOUD COMPUTING

Agenda

1 Overview

2 5 Vs Analysis

3 Architecture Plan & Data Analysis
Technologies, Services and Tools

4 Implementation & Results

Overview

- Thailand is facing flooding every year, especially during the rainy season.
- If there is no effective management, it will cause damage to people's lives, properties, and houses.
- For example, the flooding in 2024 in many areas and the reservoirs could not hold the water until they burst, such as Maha Sarakham and Lampang provinces, etc.

Objectives

- To develop a model to classify the risk of flooding in areas from rainfall data, water levels in reservoirs and dams.
- The system can display results in real time and forward decision-making recommendations to relevant agencies, such as the Royal Irrigation Department or dam officials.
- We hope this approach will help water management be more precise, reduce impacts on people, and serve as a model for using information technology to help manage water resources.

ตัวอย่าง: รูปอ่างเก็บน้ำแตกที่ จังหวัดมหาสารคาม พ.ศ. 2567

5Vs Analysis (1/2)

Volumn

There are a lot of data sources, including.

- Water level of reservoirs from the Royal Irrigation Department.
- Water level of dams from the Royal Irrigation Department.
- Risk area dataset from the Water Resources Information Institute.

Velocity

Streaming data from IoT sensors that measure water, such as rainfall and water levels in reservoirs and dams, is received in near real-time, requiring the system to be able to support rapid and continuous data flow.

Variety

The imported data comes in a variety of formats, such as a structured CSV file of the flood risk dataset and a single JSON file from the Sensor API. The semi-structured data requires data preprocessing and data transformation to format the data into a structure ready for analysis.

5Vs Analysis (2/2)

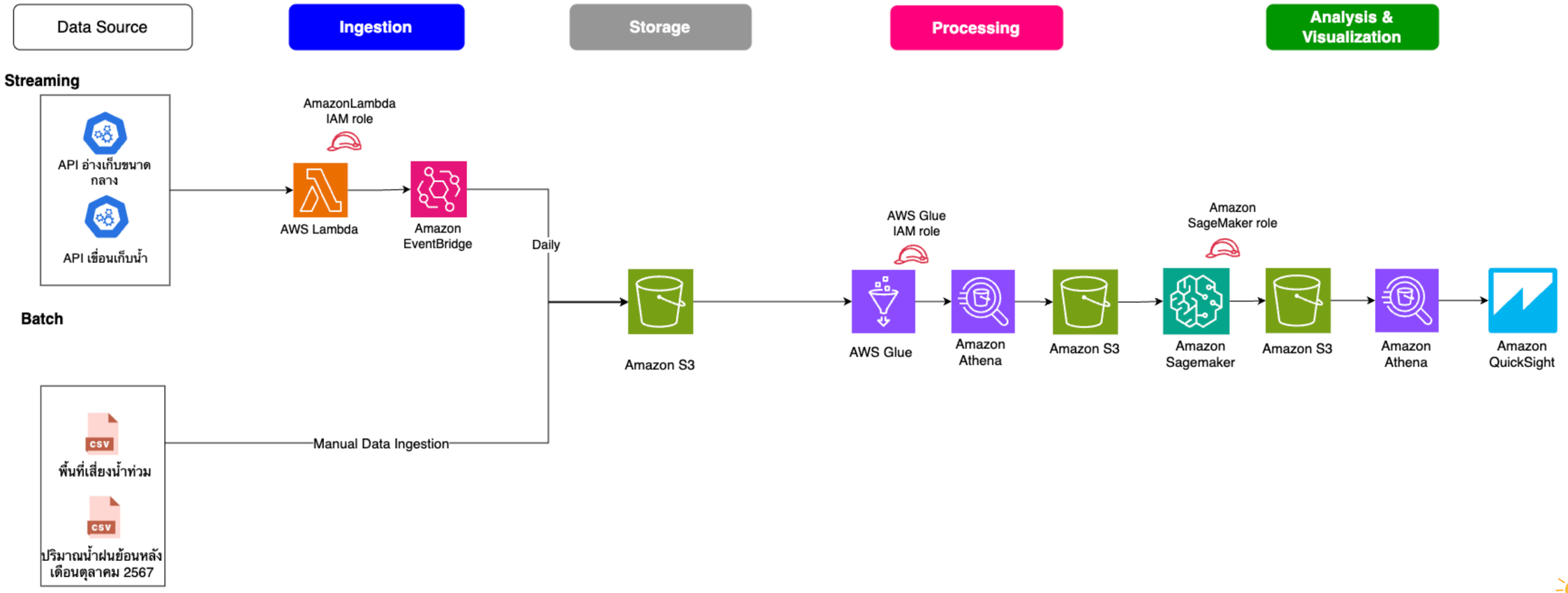
Veracity

There are steps of Data Cleaning, Outlier Detection and appropriate feature selection to reduce model errors, such as selecting only sensors with reliable data or using imputation techniques for missing data from combining multiple data sources to create training data for the model.

Value

This data, when analyzed and modeled, can help predict flood risk areas, which is very valuable for management, such as water management in dams to reduce impacts, evacuation planning, and allowing relevant agencies to make accurate and timely decisions.

Architecture Plan & Data Analysis



Technologies, Services and Tools

Data sources

Ingestion

Streaming Data Ingestion

AWS Lambda

To trigger streaming data that API Gateway received. And then, automate transfers data to store in Amazon S3.

Amazon EventBridge

To trigger Lambda function to get API information and send it to S3 bucket daily.

IAM role

To allow Lambda function put information to keep in S3 Bucket.

Batch Data Ingestion

Manual

Because the dataset of the data source is the history of dataset from the past.

Storage

Amazon S3

To store raw data from API Gateway (streaming data) and raw dataset (batch data) in S3 buckets.

Moreover, we use S3 bucket to store the data after the processing.

Processing

Streaming & Batch processing

AWS Glue

To create Glue clawler to extract, transform, and load data into an AWS Glue database. Editing schema of the table.

Moreover, creating Glue job and a trigger to run the job daily.

Amazon Athena

To preview and query data to clean, select and create the table that we need to use in the next step. .

IAM role

To allow AWS Glue access S3 Bucket and Amazon Athena.

Analysis & Visualization

Analysis

Amazon SageMaker

To use Amazon SageMaker analysis and predict the possibility of flooding areas with near-real-time.

IAM role

To allow Amazon SageMaker access S3 Bucket.

Amazon Athena

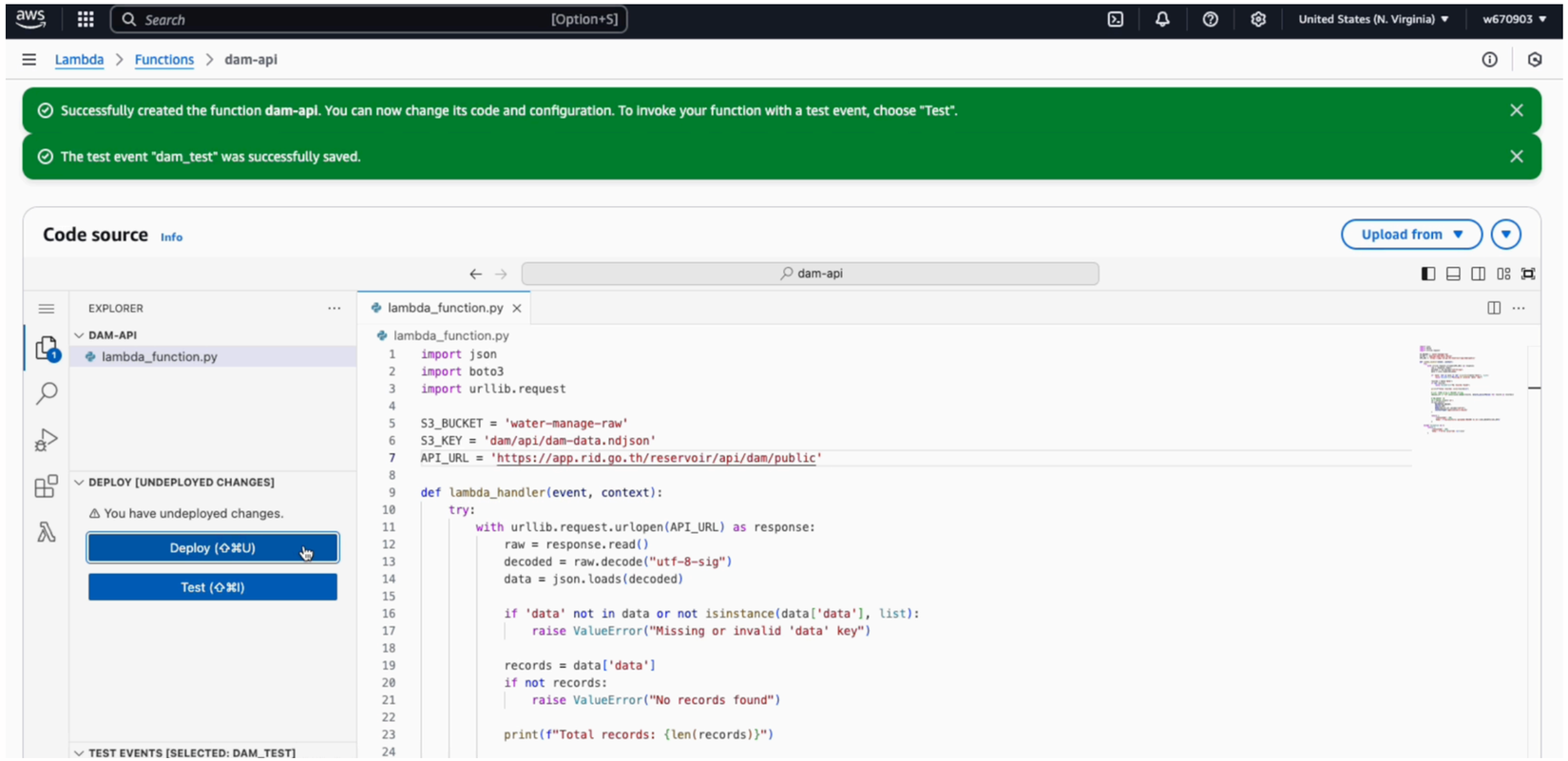
To query data and Amazon Athena pkugin .

Visualization

Amazon QuickSight

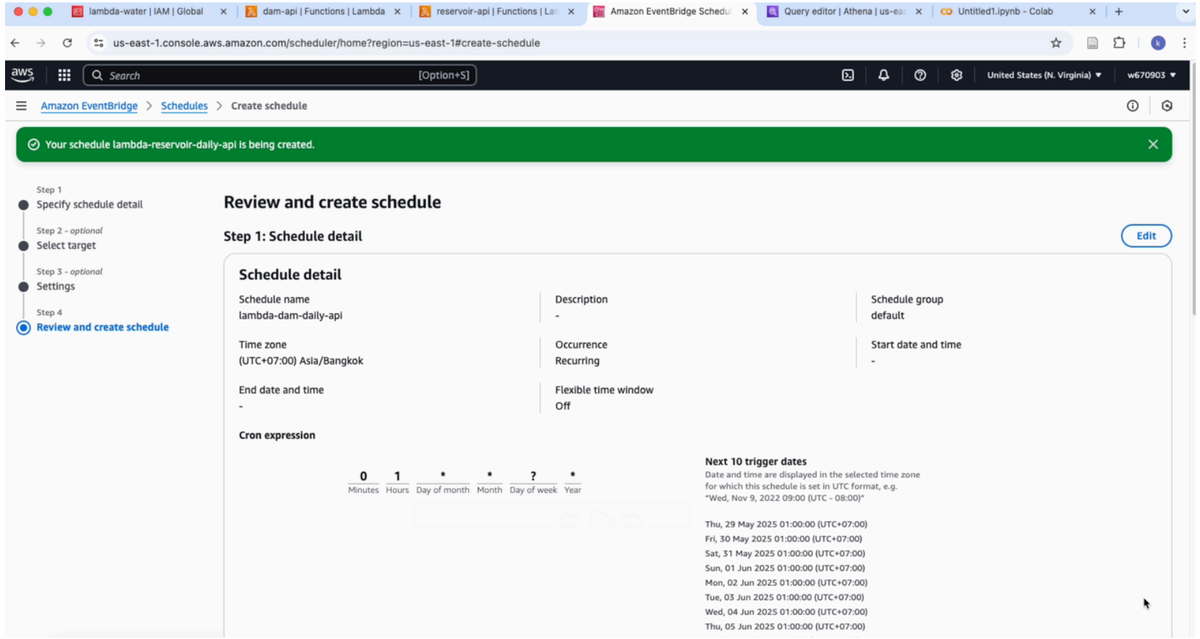
To creates a dashboard to displays the information after we analysed the cleaned data by Amazon SageMaker. We hope these data provide the advantages for related people and others who are interested in it to manage water in dam and prepare plan to handle with it before the flooding.

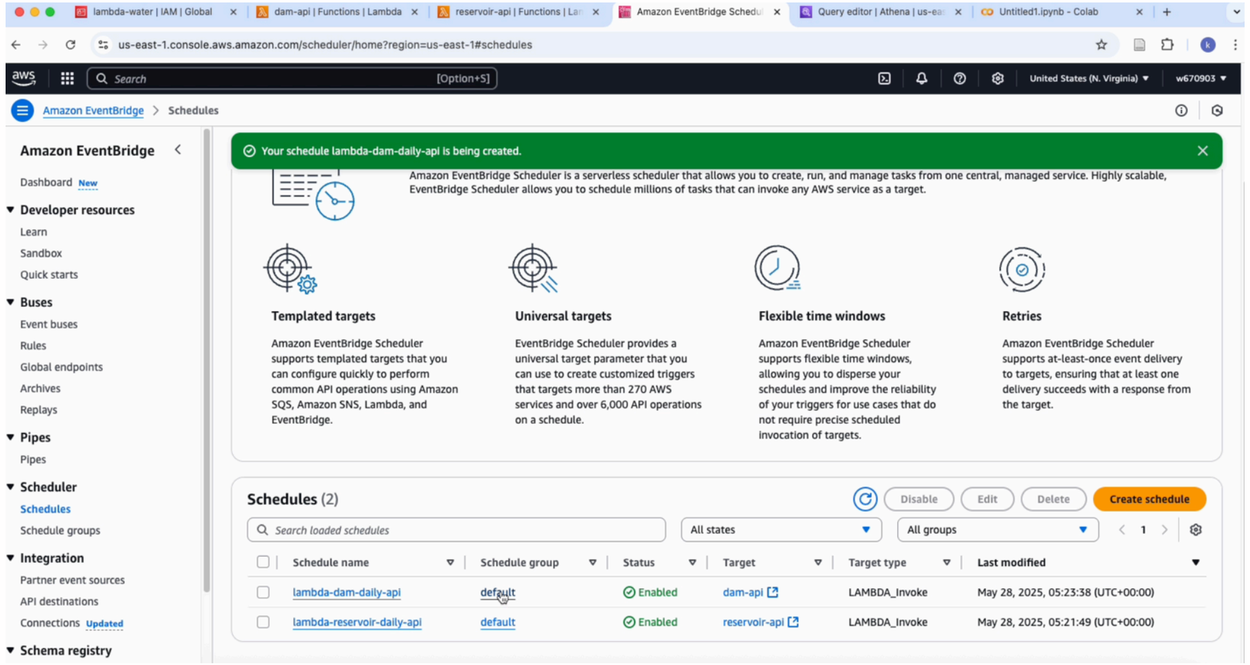
Implementation : Ingestion (1/3)



Create Lambda Function to import the water level's API

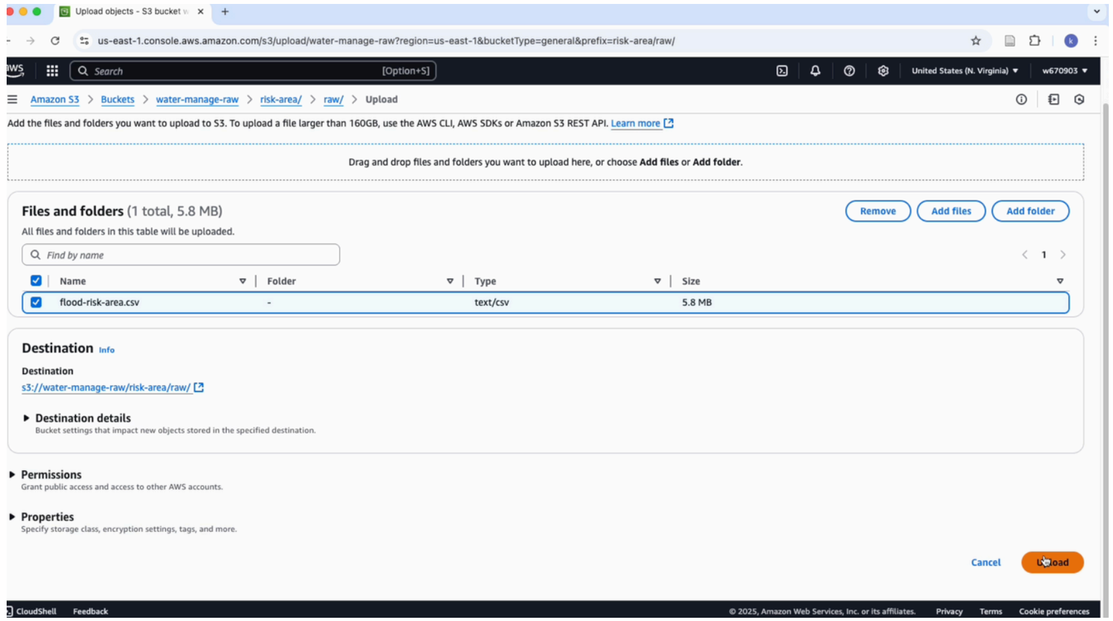
Implementation : Ingestion (2/3)

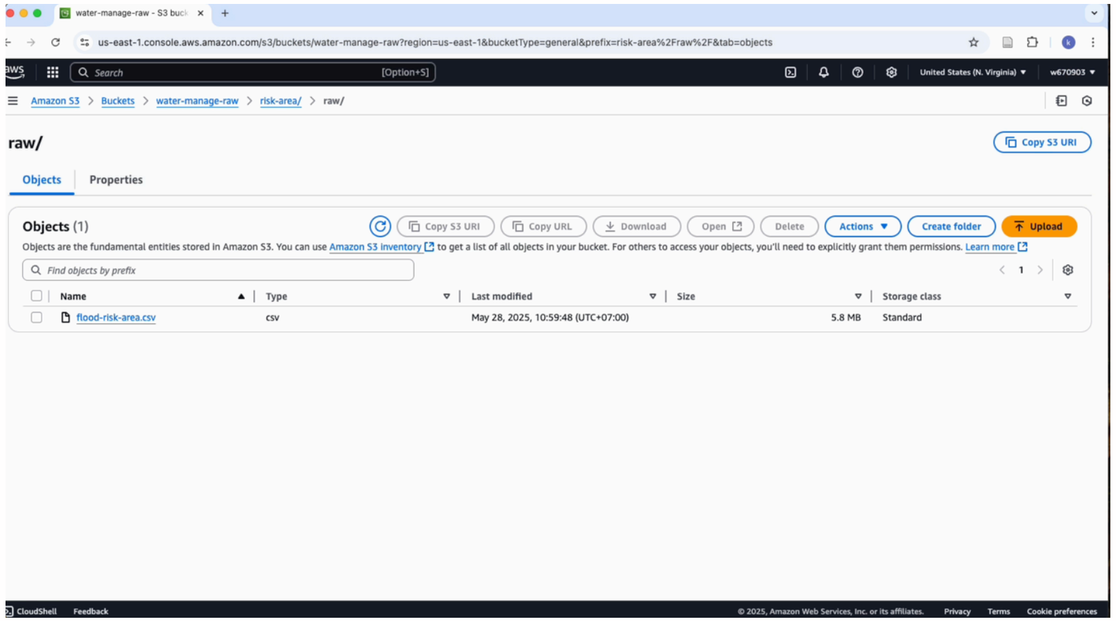




Create EventBridge Schedule to trigger Lamda Function

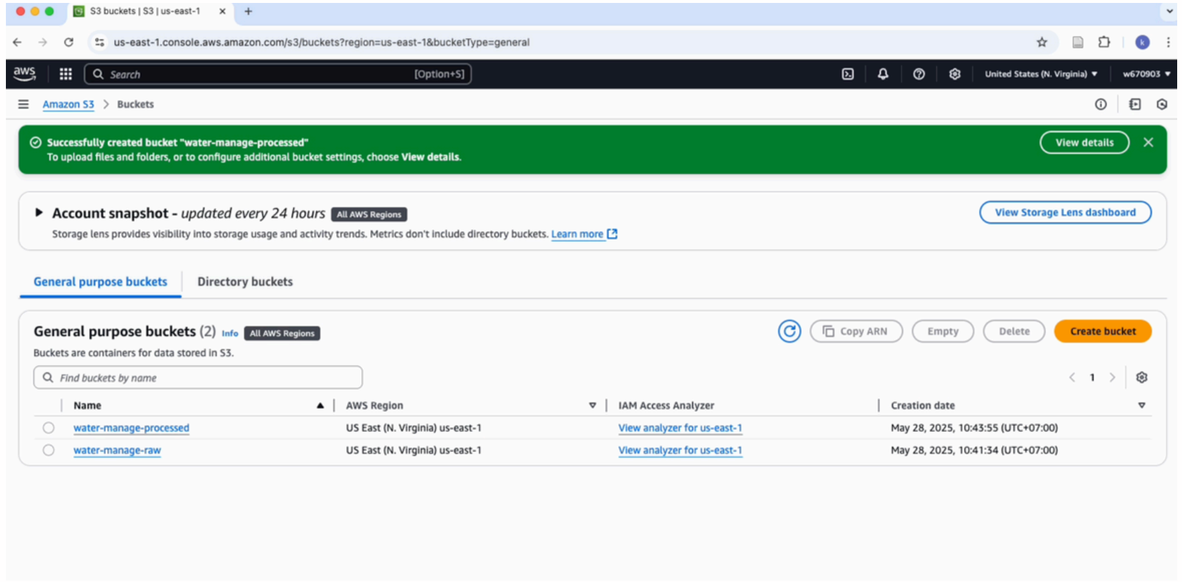
Implementation : Ingestion (3/3)

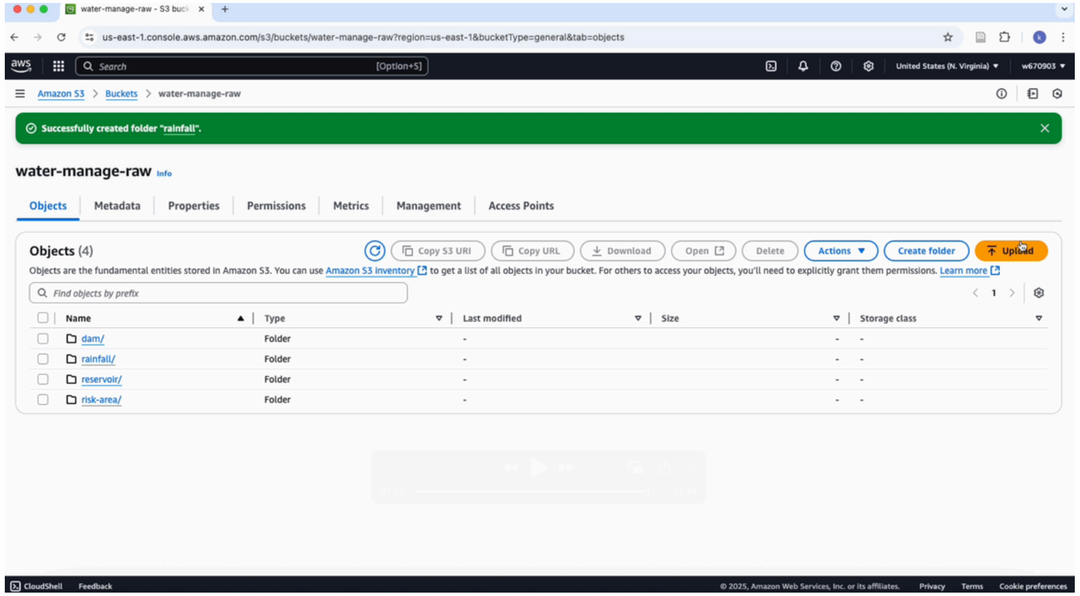




Manual upload “risk area.csv” file to S3 bucket.

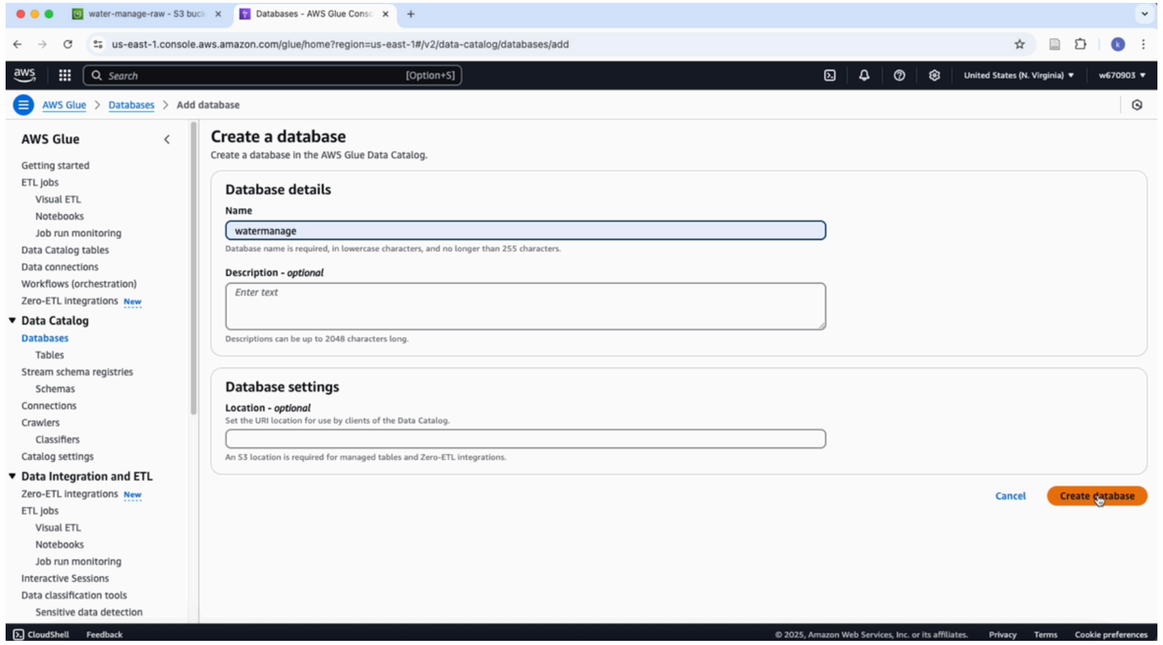
Implementation : Storage

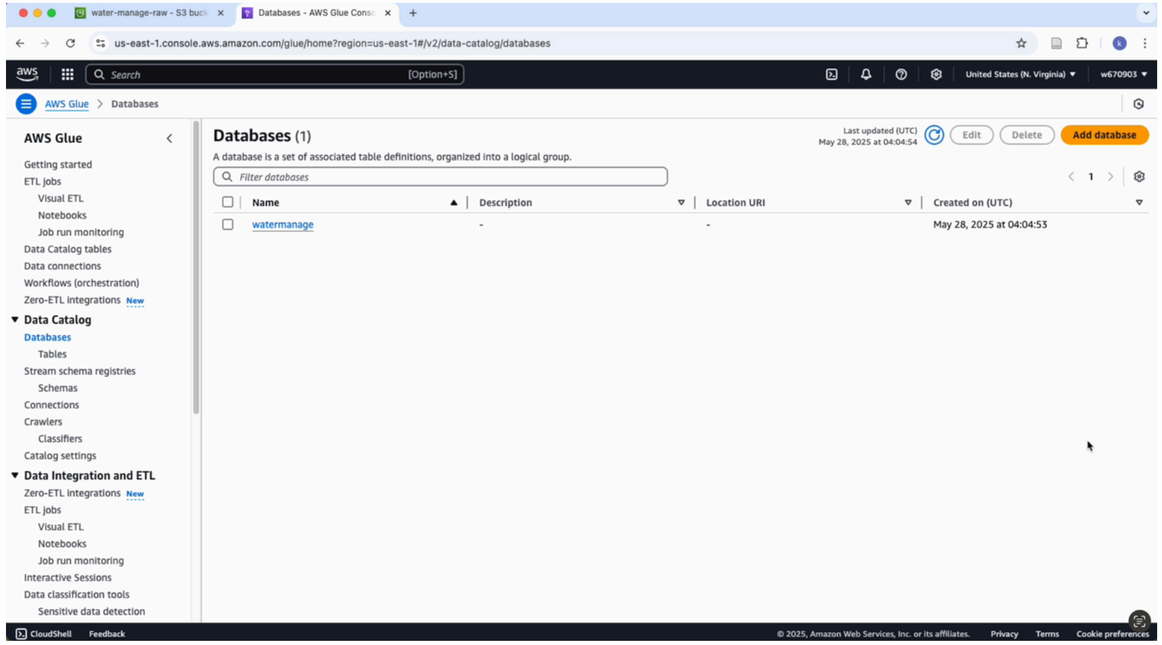




Create S3 Buckets and folders to keep raw data and processed data

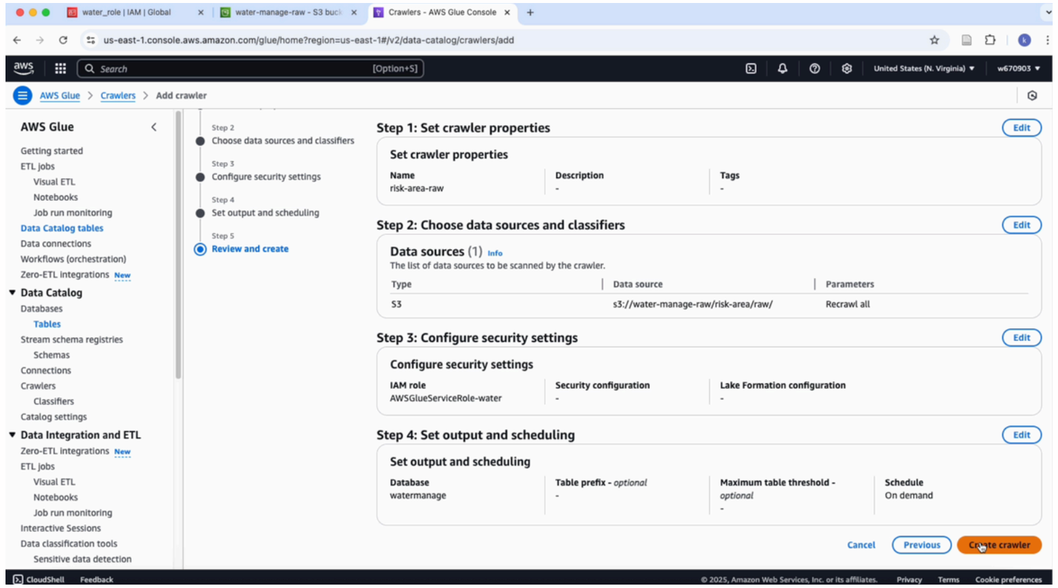
Implementation : Processing (1/4)

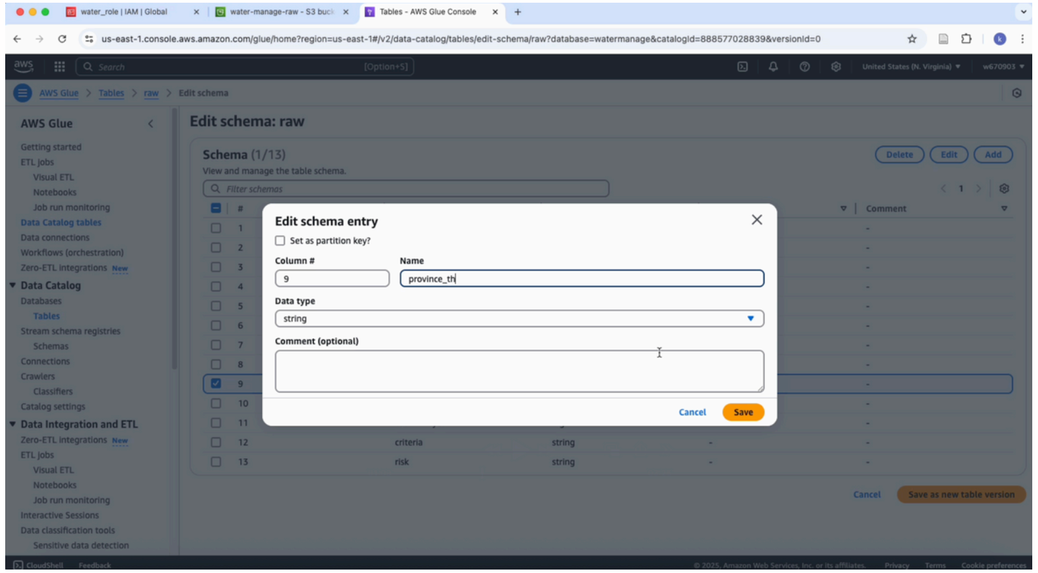




Create Glue database for supporting and managing metadata and schema.

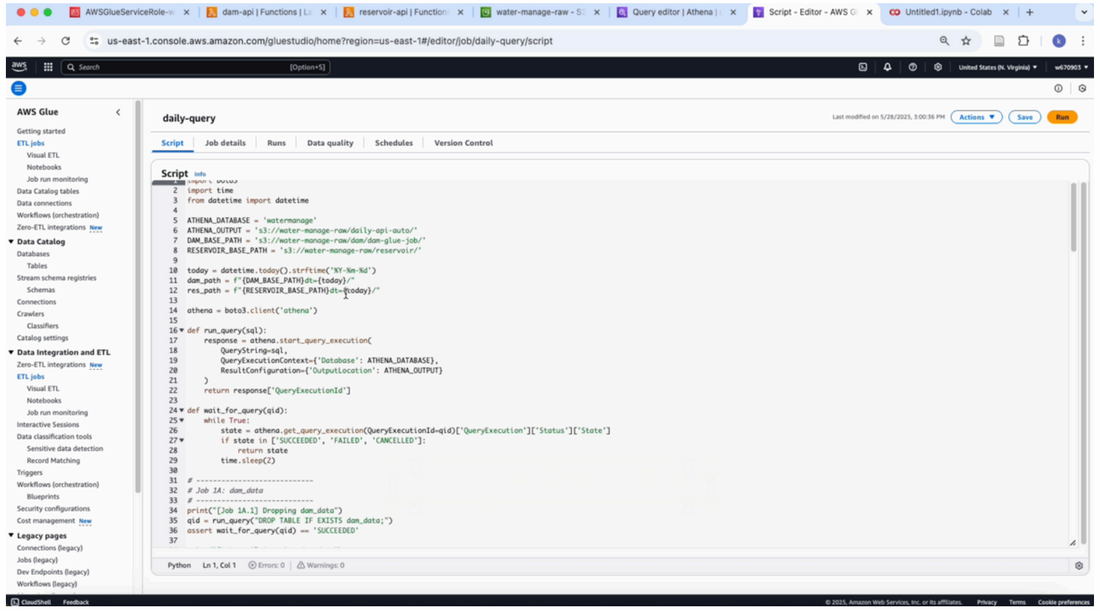
Implementation : Processing (2/4)

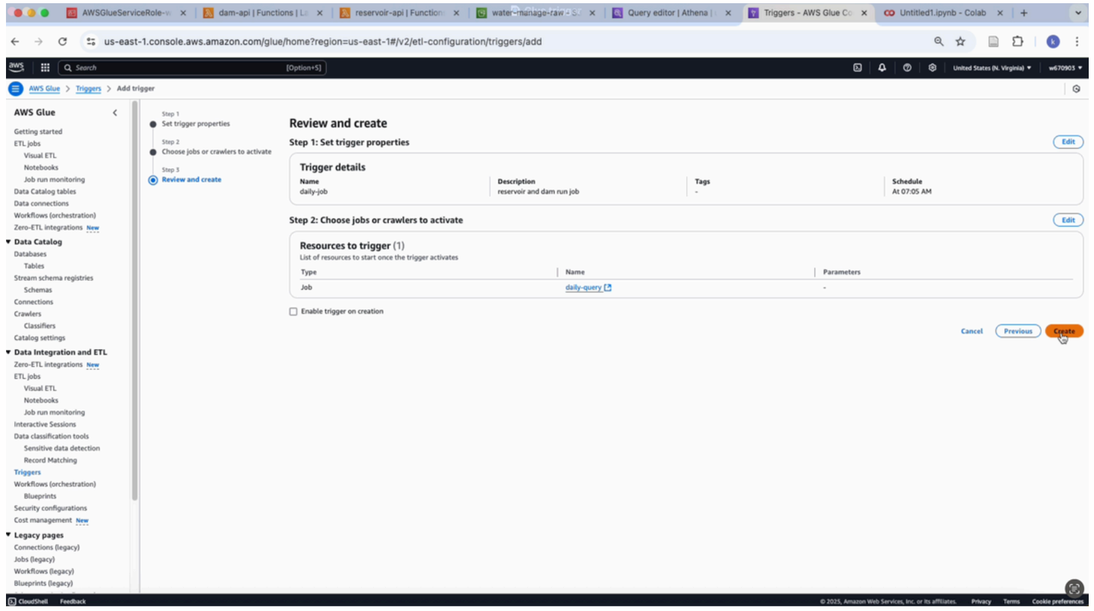




Create table by using crawler and edit schema of risk area

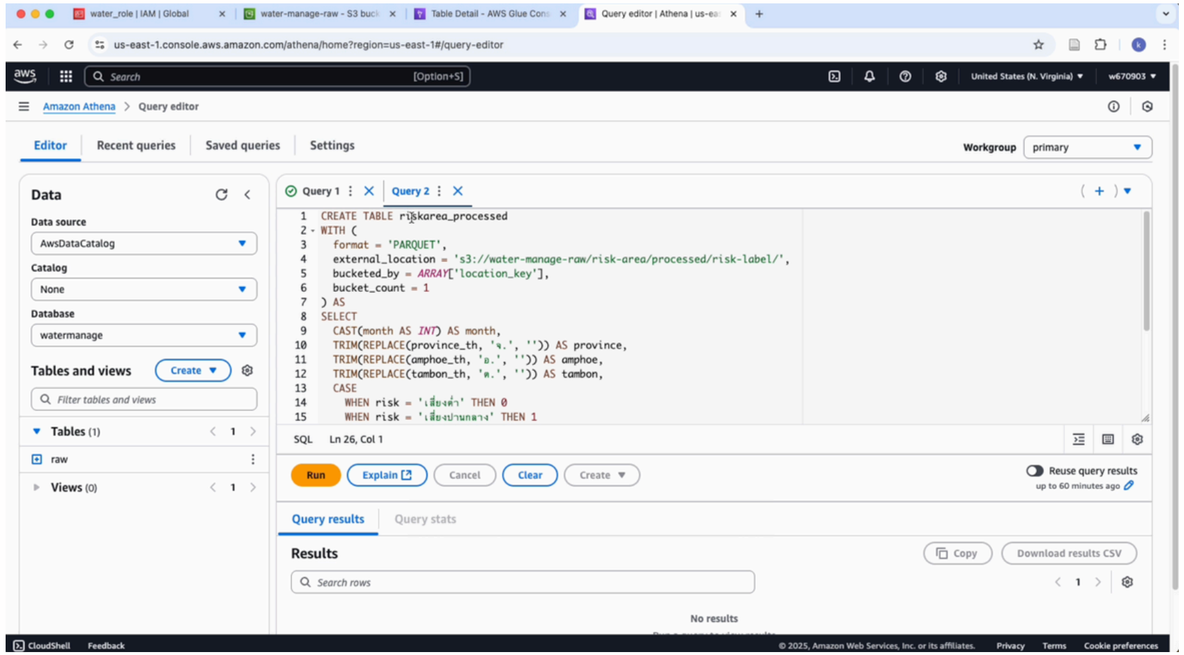
Implementation : Processing (3/4)

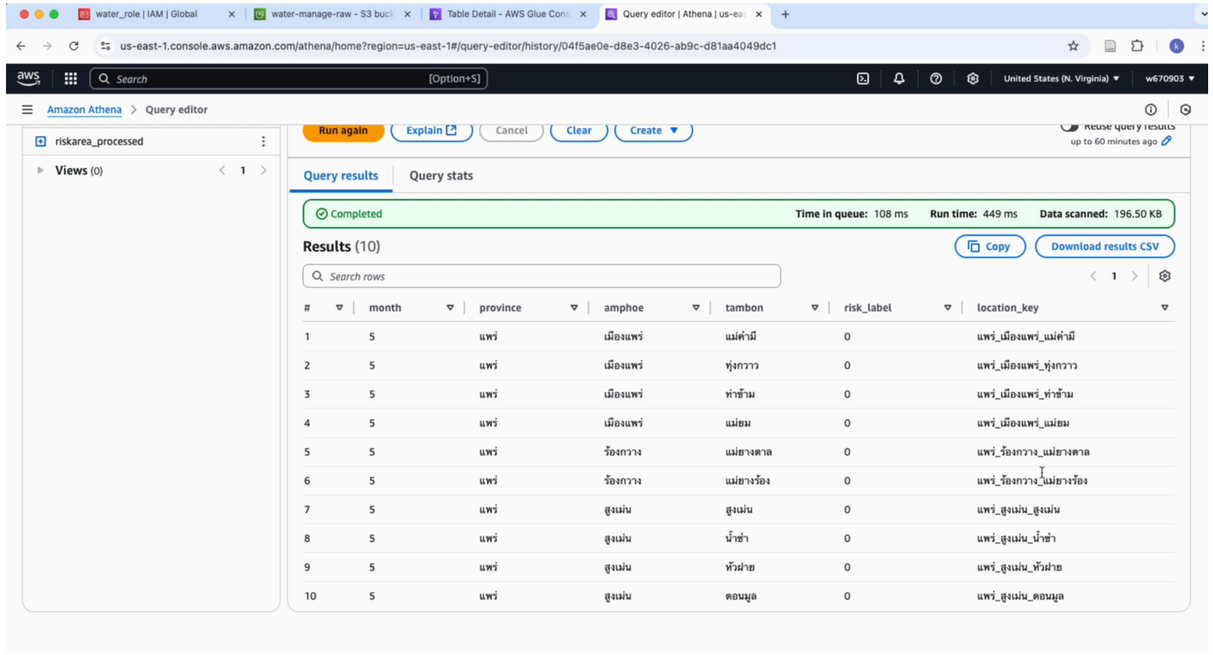




Create Glue job by using script for streaming data and add trigger for running Glue job

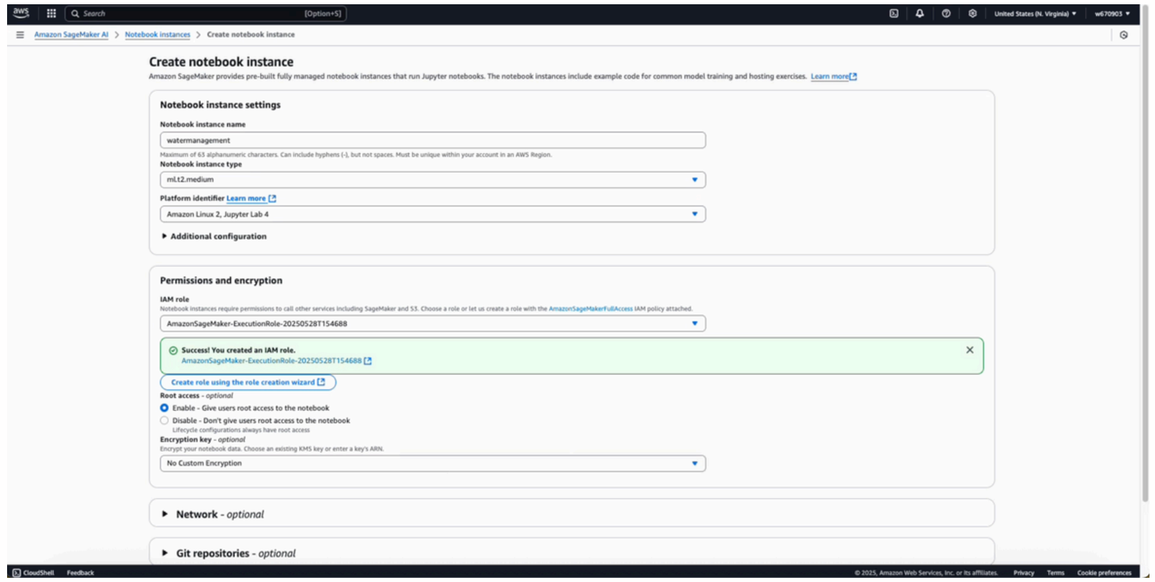
Implementation : Processing (4/4)

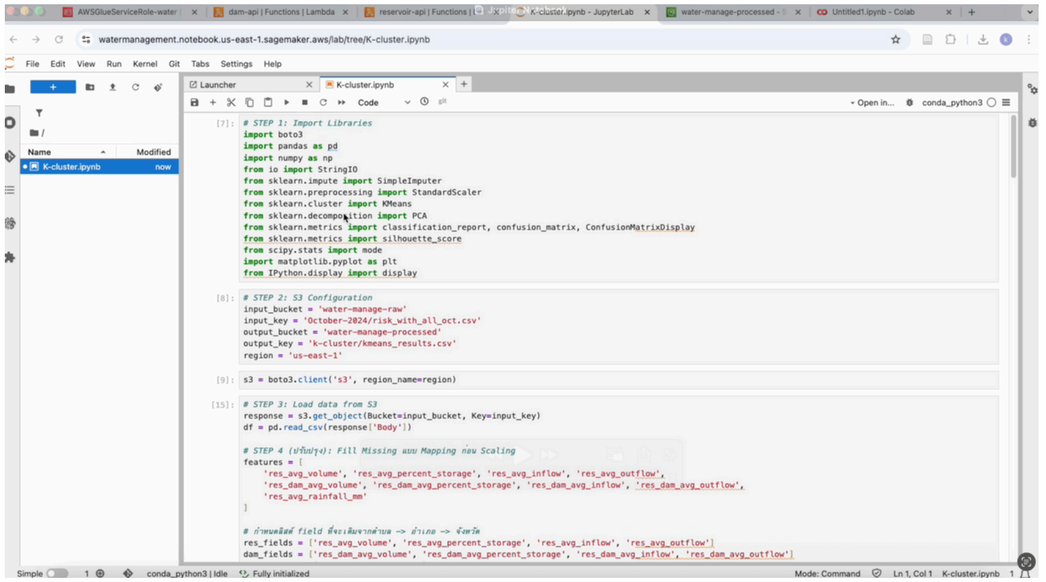




To use Amazon Athena to query and clean data

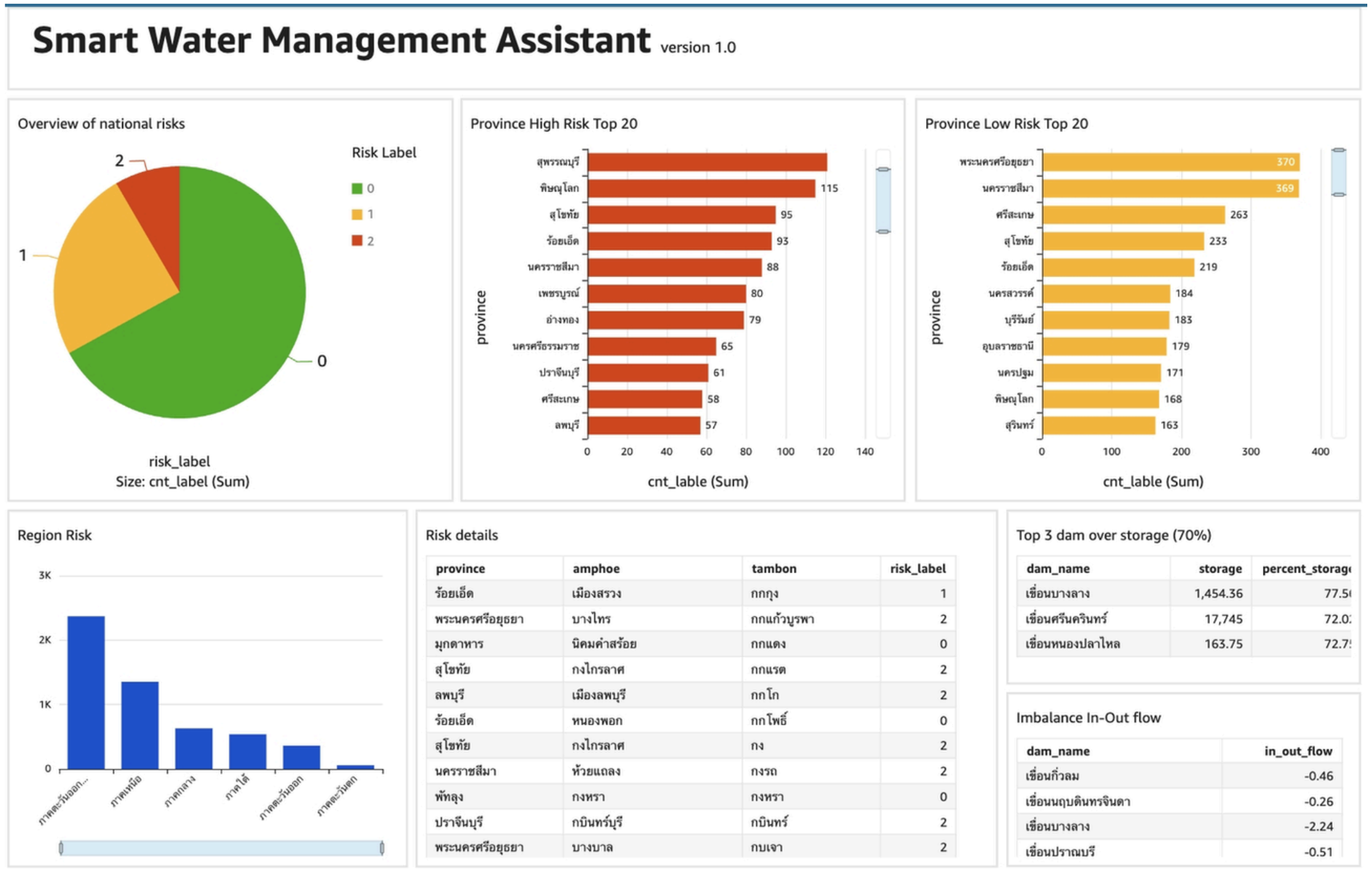
Implementation : Analysis and Visualization (1/2)





Create SageMaker Notebook and analysis data.

Implementation : Analysis and Visualization (2/2)



Visualize

Result and future work

Result

- End-to-end data pipeline developed on AWS (Lambda, Glue, Athena, SageMaker)
- Multi-source data integration (API JSON + CSV from public agencies)
- ML model predicts flood risk at sub-district level with high accuracy
- Real-time dashboard built with Grafana for spatial visualization
- Identified key challenges in data standardization and governance
- Gained practical insights on translating technical results for decision-makers

Future Work

- Integrate weather forecast data (GFS, ECMWF) for real-time prediction
- Build Early Warning System with mobile notifications (LINE/SMS)
- Apply online learning for adaptive model updates
- Improve data quality via imputation and outlier detection
- Develop interactive flood risk maps (drill-down to tambon level)
- Design user-friendly tools to support local and community-level use

APPENDIX

Milestone 2 update project progress

Data cleaning or transformation steps & How the pipeline handles data quality or noise

Batch

Glue Clawler

- Edit cloumn name for schema.

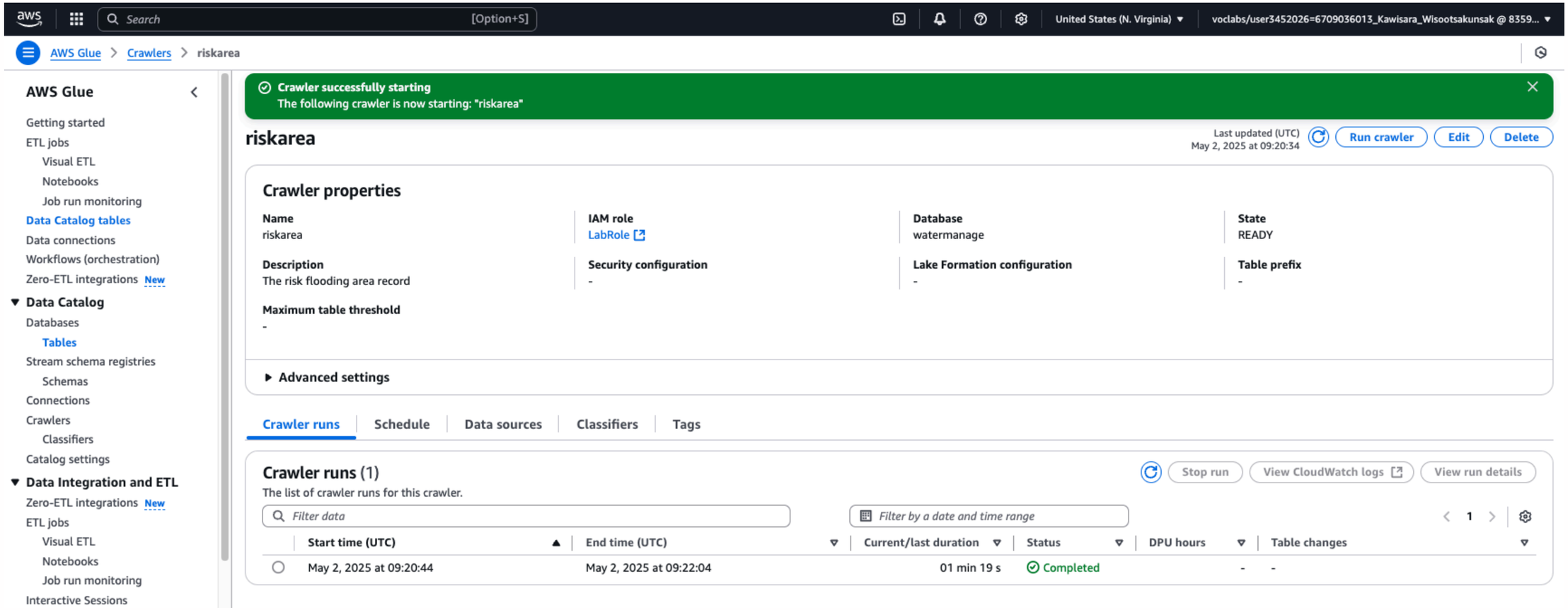
Amazon Athena

- Query data in the table (CSV file) to preview the information, i.e,. null, blank, N/A.
- Query to create the table that delete the data in the table that contain null, bank, N/A.
- Query to select columns that we would like to use in the analysis.
- Transform risk column from “string” to “integer”.

Streaming

Amazon Athena

- Query data in (CSV file) that transform from JSON (API) to preview the information, i.e,. null, blank, N/A.
- Query to create the table that delete the data in the table that contain null, bank, N/A or outlier data.
- Eliminate duplicate data.
- Query to select columns that we would like to use in the analysis.



Create Glue Clawler

Before

Schema | Partitions | Indexes | Column statistics - *new*

Schema (13)

View and manage the table schema.

Edit schema as JSON | Edit schema

Filter schemas

1

#	Column name	Data type	Partition key	Comment
1	month	bigint	-	-
2	geocode	bigint	-	-
3	tambon_t	string	-	-
4	tambon_e	string	-	-
5	amphoe_code	bigint	-	-
6	amphoe_t	string	-	-
7	amphoe_e	string	-	-
8	prov_code	bigint	-	-
9	prov_t	string	-	-
10	prov_e	string	-	-
11	count 17 year	bigint	-	-
12	criteria	string	-	-
13	risk	string	-	-

After

Schema | Partitions | Indexes | Column statistics - *new*

Schema (13)

View and manage the table schema.

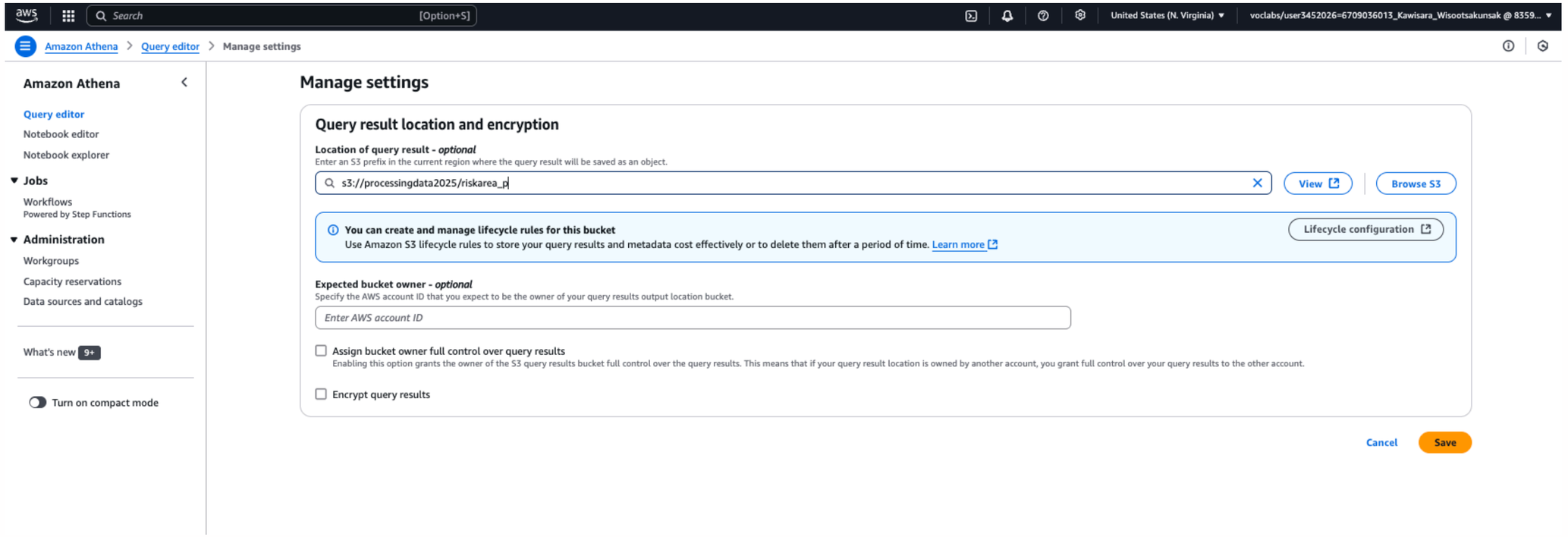
Edit schema as JSON | Edit schema

Filter schemas

1

#	Column name	Data type	Partition key	Comment
1	month	bigint	-	-
2	geocode	bigint	-	-
3	sub_district_th	string	-	-
4	sub_district_en	string	-	-
5	district_code	bigint	-	-
6	district_th	string	-	-
7	district_en	string	-	-
8	province_code	bigint	-	-
9	province_th	string	-	-
10	province_eh	string	-	-
11	count_17_years	bigint	-	-
12	criteria	string	-	-
13	risk	string	-	-

Edit column name and create schema



Select S3 bucket to store query table

Query results | Query stats

Completed — Time in queue: 107 ms — Run time: 677 ms — Data scanned: 816.90 KB

Results (10)

Copy | Download results CSV

Search rows

1

#	month	geocode	sub_district_th	sub_district_en	district_code	district_th	district_en	province_code	province_th
1	1	120501	ต.ไทรน้อย	Sai Noi	1205	อ.ไทรน้อย	Sai Noi District	12	จ.นนทบุรี
2	1	120503	ต.หนองเพรางาย	Nong Phrao Ngai	1205	อ.ไทรน้อย	Sai Noi District	12	จ.นนทบุรี
3	1	120504	ต.ไทรใหญ่	Sai Yai	1205	อ.ไทรน้อย	Sai Noi District	12	จ.นนทบุรี
4	1	120505	ต.ขุนศรี	Khun Si	1205	อ.ไทรน้อย	Sai Noi District	12	จ.นนทบุรี
5	1	120507	ต.ทวีวัฒนา	Thawi Watthana	1205	อ.ไทรน้อย	Sai Noi District	12	จ.นนทบุรี
6	1	140110	ต.สวนพริก	Suan Phrik	1401	อ.พระนครศรีอยุธยา	Phra Nakhon Si Ayutthaya District	14	จ.พระนครศรีอยุธยา
7	1	140507	ต.บ้านคลัง	Ban Khlang	1405	อ.บางบาล	Bang Ban District	14	จ.พระนครศรีอยุธยา
8	1	140509	ต.น้ำเต้า	Namtao	1405	อ.บางบาล	Bang Ban District	14	จ.พระนครศรีอยุธยา
9	1	140511	ต.วัดตะกู	Wat Taku	1405	อ.บางบาล	Bang Ban District	14	จ.พระนครศรีอยุธยา
10	1	140512	ต.บางหลวง	Bang Luang	1405	อ.บางบาล	Bang Ban District	14	จ.พระนครศรีอยุธยา

Preview data in table

Query results | Query stats

Completed — Time in queue: 121 ms — Run time: 514 ms — Data scanned: 115.40 KB

Results (10)

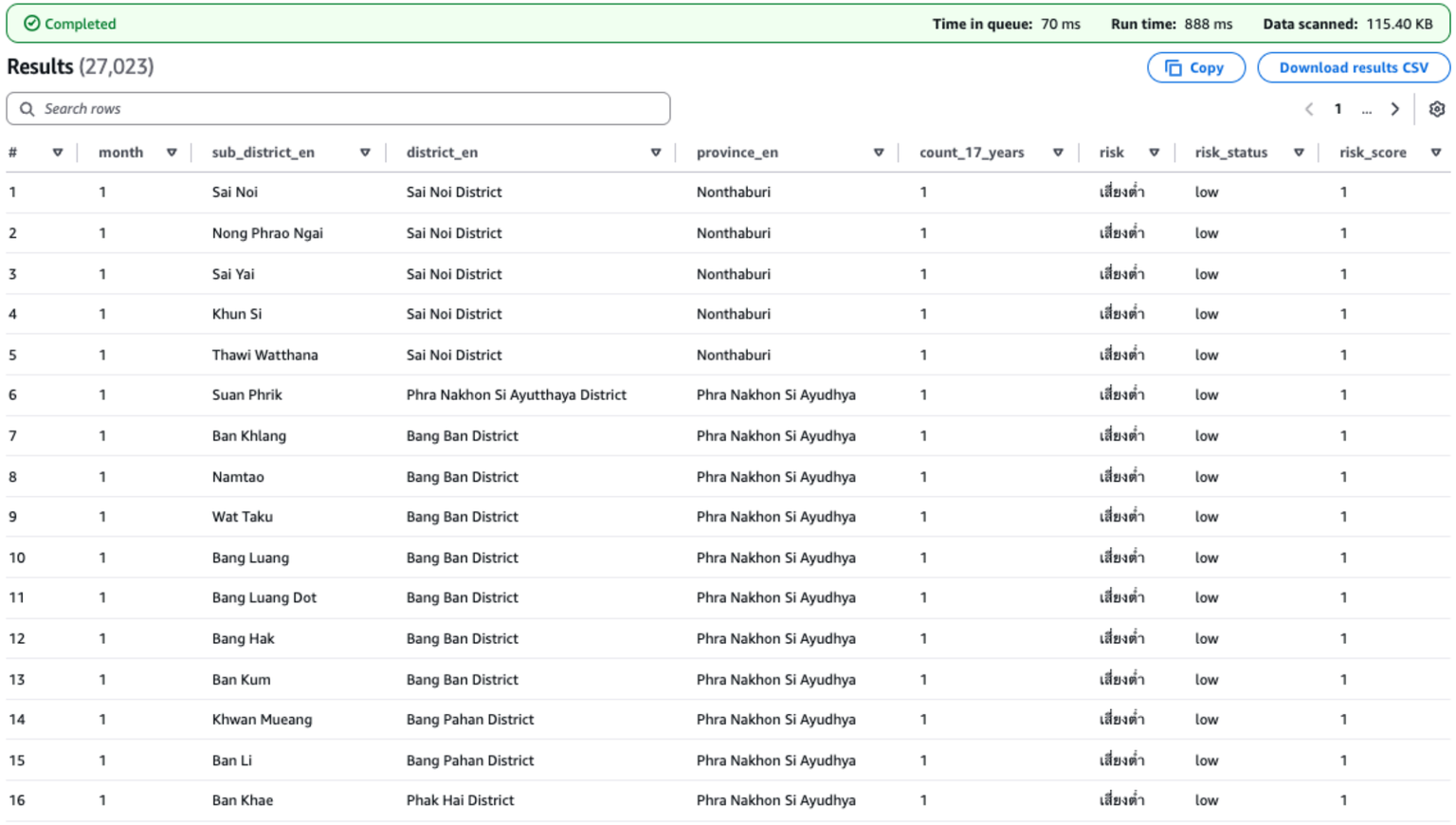
Copy | Download results CSV

Search rows

1

#	month	sub_district_en	district_en	province_en	count_17_years	risk
1	1	Sai Noi	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ
2	1	Nong Phrao Ngai	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ
3	1	Sai Yai	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ
4	1	Khun Si	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ
5	1	Thawi Watthana	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ
6	1	Suan Phrik	Phra Nakhon Si Ayutthaya District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ
7	1	Ban Khlang	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ
8	1	Namtao	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ
9	1	Wat Taku	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ
10	1	Bang Luang	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ

Only Select table that would like to use



Completed — Time in queue: 70 ms — Run time: 888 ms — Data scanned: 115.40 KB

Results (27,023)

Copy | Download results CSV

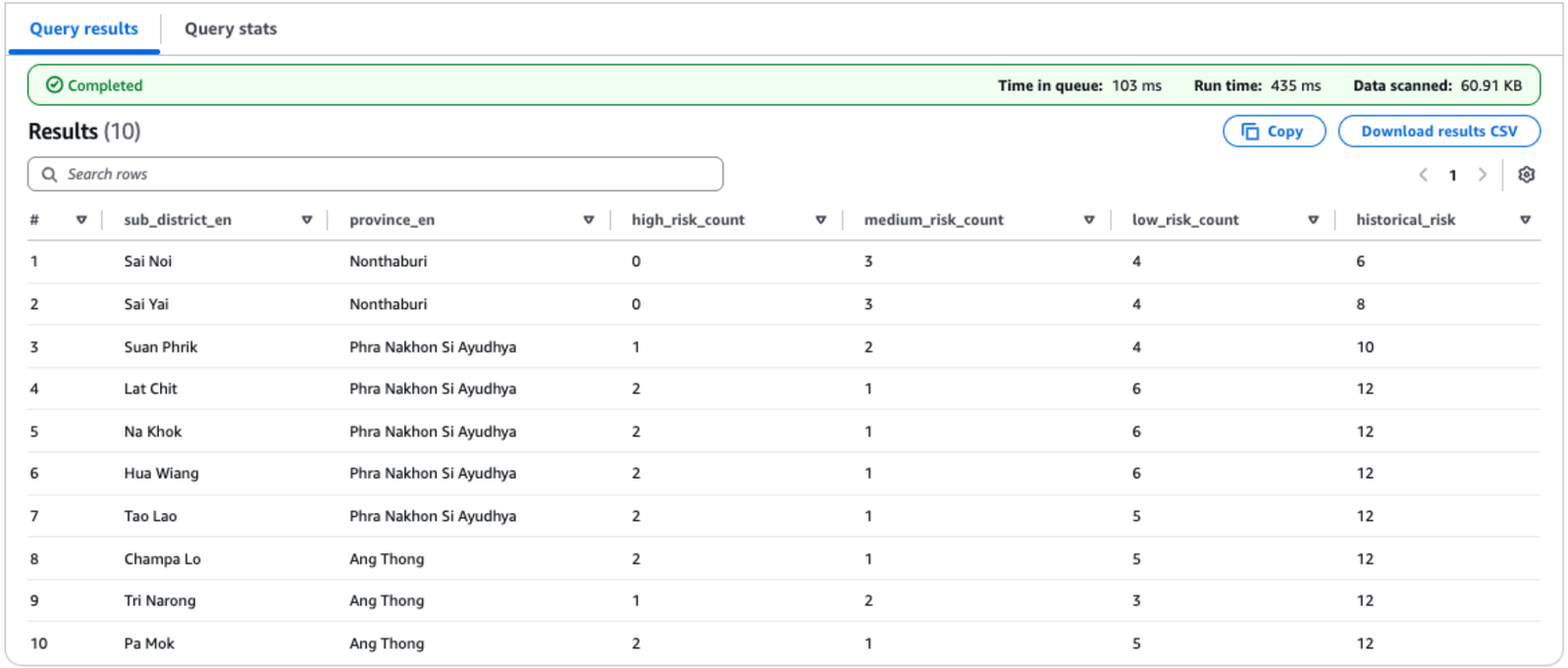
Search rows

1 ...

#	month	sub_district_en	district_en	province_en	count_17_years	risk	risk_status	risk_score
1	1	Sai Noi	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ	low	1
2	1	Nong Phrao Ngai	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ	low	1
3	1	Sai Yai	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ	low	1
4	1	Khun Si	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ	low	1
5	1	Thawi Watthana	Sai Noi District	Nonthaburi	1	เสี่ยงต่ำ	low	1
6	1	Suan Phrik	Phra Nakhon Si Ayutthaya District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
7	1	Ban Khlang	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
8	1	Namtao	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
9	1	Wat Taku	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
10	1	Bang Luang	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
11	1	Bang Luang Dot	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
12	1	Bang Hak	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
13	1	Ban Kum	Bang Ban District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
14	1	Khwan Mueang	Bang Pahan District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
15	1	Ban Li	Bang Pahan District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1
16	1	Ban Khae	Phak Hai District	Phra Nakhon Si Ayudhya	1	เสี่ยงต่ำ	low	1

Query to transform data type

Batch : Screenshots or tables of sample processed data



Query results | Query stats

Completed — Time in queue: 103 ms — Run time: 435 ms — Data scanned: 60.91 KB

Results (10)

Copy | Download results CSV

Search rows

1

#	sub_district_en	province_en	high_risk_count	medium_risk_count	low_risk_count	historical_risk
1	Sai Noi	Nonthaburi	0	3	4	6
2	Sai Yai	Nonthaburi	0	3	4	8
3	Suan Phrik	Phra Nakhon Si Ayudhya	1	2	4	10
4	Lat Chit	Phra Nakhon Si Ayudhya	2	1	6	12
5	Na Khok	Phra Nakhon Si Ayudhya	2	1	6	12
6	Hua Wiang	Phra Nakhon Si Ayudhya	2	1	6	12
7	Tao Lao	Phra Nakhon Si Ayudhya	2	1	5	12
8	Champa Lo	Ang Thong	2	1	5	12
9	Tri Narong	Ang Thong	1	2	3	12
10	Pa Mok	Ang Thong	2	1	5	12

The processed data of risking flooding area for Risk Clustering Analysis.

Batch : Screenshots or tables of sample processed data

Query results | Query stats

Completed — Time in queue: 70 ms — Run time: 527 ms — Data scanned: 137.60 KB

Results (10)

Copy | Download results CSV

Search rows

1

#	month	province_en	district_en	sub_district_en	avg_risk_score
1	6	Loei	Tha Li District	A Hi	1.0
2	2	Yala	Raman District	A Song	1.0
3	11	Yala	Raman District	A Song	1.0
4	12	Yala	Raman District	A Song	3.0
5	7	Loei	Tha Li District	A Hi	1.0
6	11	Pattani	Mueang Pattani District	A Noru	1.0
7	12	Pattani	Mueang Pattani District	A Noru	1.0
8	7	Surin	Buachet District	A Phon	1.0
9	9	Surin	Buachet District	A Phon	1.0
10	1	Yala	Raman District	A Song	2.0

The processed data of average risking flooding score for Monthly Risk Trend Analysis.

Streaming : Information for processing data


```
{
    "document": "https://app.rid.go.th/reservoir/api/document/dam",
    "date": "2025-03-27",
    "total": 35,
    "data": [
        {
            "region": "ภาคเหนือ",
            "dam": [
                {
                    "id": "100104",
                    "name": "เขื่อนแม่กวงอุดมธารา",
                    "owner": "กรมชลประทาน",
                    "capacity": 295,
                    "storage": 263,
                    "active_storage": 249,
                    "dead_storage": 14,
                    "volume": 151.61,
                    "percent_storage": 57.65,
                    "inflow": 0.13,
                    "outflow": 0.62
                },
                {
                    "id": "100105",
                    "name": "เขื่อนกิ่วลม",
                    "owner": "กรมชลประทาน",
                    "capacity": 106.22,
                    "storage": 106.22,
                    "active_storage": 102.67,
                    "dead_storage": 3.55,
                    "volume": 46.89,
                    "percent_storage": 44.14,
                    "inflow": 0.07,
                    "outflow": 2.24
                },
```

```
"measurementResults": [
  {
    "measureTime": "2025-05-02T07:00:00",
    "createTime": "2025-05-02T07:00:33",
    "updateTime": "2025-05-02T07:00:33",
    "variable": "Rainfall",
    "value": 0.5,
    "uom": "mm",
    "qualityFlag": "U",
    "comment": "No quality control",
    "qualityControlLevel": "1"
  }
]
},
```


The processed data of reservoir and rainfalls for **Risk Analysis of Reservoir Overcapacity.**

Streaming : Information for processing data


```
{
    "document": "https://app.rid.go.th/reservoir/api/document/dam",
    "date": "2025-03-27",
    "total": 35,
    "data": [
        {
            "region": "ภาคเหนือ",
            "dam": [
                {
                    "id": "100104",
                    "name": "เขื่อนแม่กวงอุดมธารา",
                    "owner": "กรมชลประทาน",
                    "capacity": 295,
                    "storage": 263,
                    "active_storage": 249,
                    "dead_storage": 14,
                    "volume": 151.61,
                    "percent_storage": 57.65,
                    "inflow": 0.13,
                    "outflow": 0.62
                },
                {
                    "id": "100105",
                    "name": "เขื่อนกิ่วลม",
                    "owner": "กรมชลประทาน",
                    "capacity": 106.22,
                    "storage": 106.22,
                    "active_storage": 102.67,
                    "dead_storage": 3.55,
                    "volume": 46.89,
                    "percent_storage": 44.14,
                    "inflow": 0.07,
                    "outflow": 2.24
                },
```


The processed data of reservoir for Forecasting of Reservoir Water Discharge.

Streaming : Information for processing data


```
{
    "document": "https://app.rid.go.th/reservoir/api/document/dam",
    "date": "2025-03-27",
    "total": 35,
    "data": [
        {
            "region": "ภาคเหนือ",
            "dam": [
                {
                    "id": "100104",
                    "name": "เขื่อนแม่กวงอุดมธารา",
                    "owner": "กรมชลประทาน",
                    "capacity": 295,
                    "storage": 263,
                    "active_storage": 249,
                    "dead_storage": 14,
                    "volume": 151.61,
                    "percent_storage": 57.65,
                    "inflow": 0.13,
                    "outflow": 0.62
                },
                {
                    "id": "100105",
                    "name": "เขื่อนกิ่วลม",
                    "owner": "กรมชลประทาน",
                    "capacity": 106.22,
                    "storage": 106.22,
                    "active_storage": 102.67,
                    "dead_storage": 3.55,
                    "volume": 46.89,
                    "percent_storage": 44.14,
                    "inflow": 0.07,
                    "outflow": 2.24
                },
```

```
"measurementResults": [
  {
    "measureTime": "2025-05-02T07:00:00",
    "createTime": "2025-05-02T07:00:33",
    "updateTime": "2025-05-02T07:00:33",
    "variable": "Rainfall",
    "value": 0.5,
    "uom": "mm",
    "qualityFlag": "U",
    "comment": "No quality control",
    "qualityControlLevel": "1"
  }
]
},
```


The processed data of reservoir and rainfalls for **Daily Change Analysis.**